Computers Chem. Vol. 14, No. 1, pp. 49–57, 1990
Printed in Great Britain

0097-8485/90 $3.00 + 0.00
Pergamon Press plc

A CRITICAL COMPARISON OF LEAST ABSOLUTE DEVIATION FITTING (ROBUST) AND LEAST SQUARES FITTING: THE IMPORTANCE OF ERROR DISTRIBUTIONS

I. B. C. MATHESON
On Line Instrument Systems, P. O. Box 111, Route 2, Jefferson, GA 30549, U.S.A.

(*Received* 2 *September* 1988; *received for publication* 19 *January* 1989)

Abstract—Non-linear least absolute deviation fitting of data has been shown to be superior to least squares where the data errors are unevenly distributed about the function. The methods give insignificantly different results for evenly-distributed errors. Criteria are given for choosing between least absolute deviation and least squares error minimization. Optical density measurements have non-Gaussian error distributions and are better fitted with minimization of the absolute deviation.

INTRODUCTION

A simple method for the robust non-linear fitting of any function has been recently developed (Matheson, 1989). The method assumes that the errors associated with the data may be represented by a double-sided exponential error distribution. It minimizes the sum of the absolute values of the deviations, the deviations being the differences between the experimental data and the calculated fit.

The non-linear fitting procedure used was that of Levenberg (1944) and Marquardt (1963), using a program form derived from that of Bevington (1969). The Levenberg–Marquardt method (LM) uses a combination of linearization of the function and steepest descent to achieve convergence, and is much favoured by experimentalists, being relatively immune to finding false minima. Parameter values, either input directly or obtained by an approximation method, are used to calculate the function and the derivations of each parameter with respect to function for all data pairs *I*. The residual for each *I*, *Y*RES, is calculated as *Y*DATA − *Y*FIT. In the case of non-linear least squares, NLLS, a β vector is calculated for each parameter as the product of the derivative times the residual and an α array as a product to two suitably permutated derivatives. The core of the LM method is the generation of adjustments, ADJ, by solution of the matrix equation:

$$\mathrm{ADJ} = \beta/\alpha. \qquad (1)$$

The non-linear least absolute deviation method, NLLAD, still solves equation (1) but uses differently derived β and α. It has been shown (1), on the basis of maximum likelihood criteria of Huber (1977), that the modification required for the vector β and the matrix α is simply division by the absolute value of the residual for each *I*, i.e.

$$\beta = \mathrm{DERIV}^{*}\,Y\mathrm{RES}/\mathrm{abs}(Y\mathrm{RES}),$$

$$\alpha = \alpha/\mathrm{abs}(YRES).$$

The crucial difference between NLLAD and NLLS is that for NLLS, α and by extension ADJ are proportional to the residual *Y*RES whereas for NLLAD they are proportional only to the sign of *Y*RES. Thus the adjustments and the final converged values of the parameters are much less influenced by outlying points for NLLAD than for NLLS. This resistance to outliers confers "robustness" on NLLAD.

The previous paper (Matheson, 1989) used literature data sets that were relatively sparse, 10–100 data pairs. The purpose of the present work is to extend the observations to larger data sets and to compare critically NLLAD to NLLS. The applicability of robust fitting, NLLAD, in circumstances highly unfavourable to NLLS will be indicated.

RESULTS

Synthetic data

The advantage of synthetic data is that the error distribution can be defined precisely and a successful fit should always return the parameter values used for synthesis within statistical error. Three types of noise were considered, Gaussian, Poissonian and double-sided exponential (DSX). The Gaussian, also known as the normal distribution, is applicable to a large number of events with a small success probability (Bevington, 1969). The Gaussian distribution is explicity assumed for the least squares method. The Poissonian distribution is more applicable to a small number of events with a large success probability and is commonly assumed for counting experiments. A Poissonian distribution is approximated by multiplication of Gaussian residuals by the square-root of the amplitude for each point. The DSX distribution is much more sharply peaked and is more extensive in the wings than the normal distribution. The DSX distribution was explicitly assumed in the derivation of NLLAD.

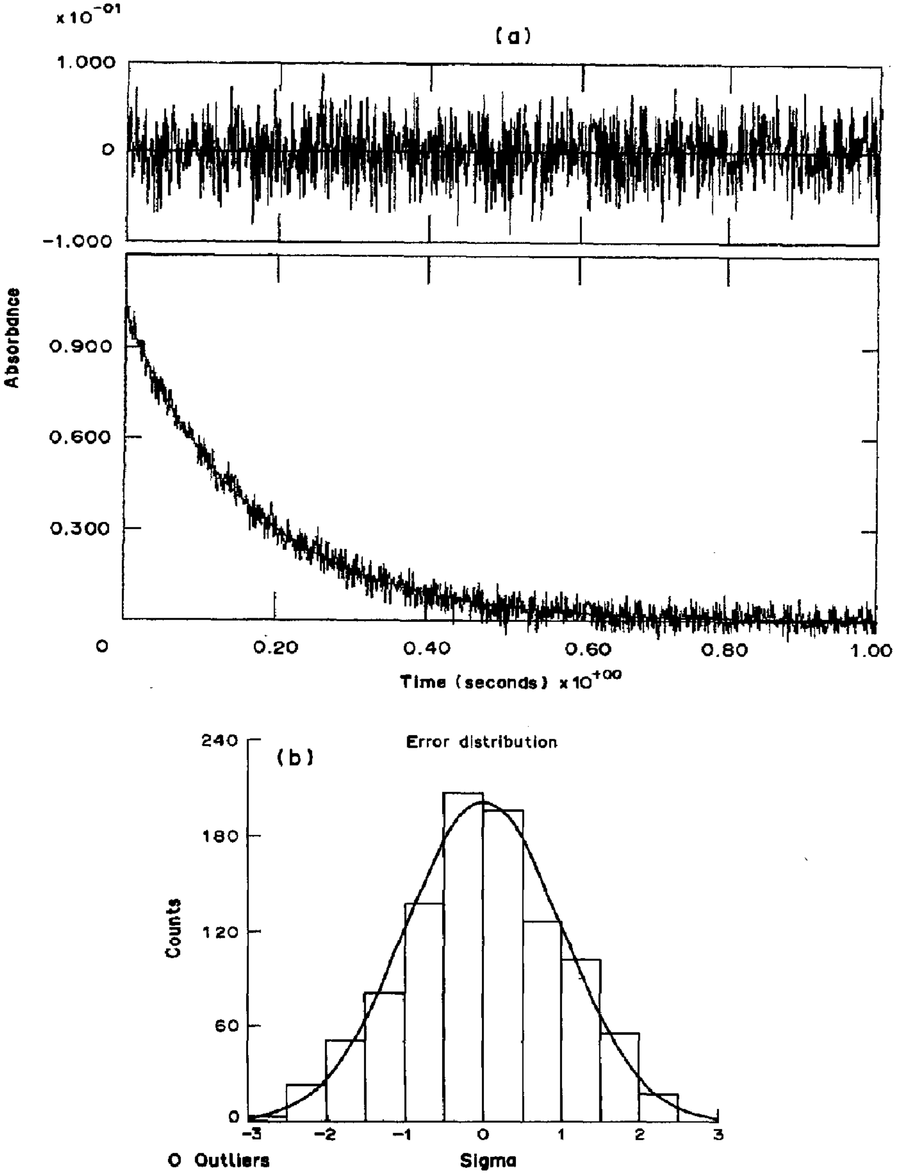


Fig. 1. Synthetic Gaussian data. Optical density as a function of time (a); histogram of residuals (b).

Data corresponding to a first-order decay were synthesized as 1000 points with a rate constant of $6\,s^{-1}$ and an amplitude of 1. 0.03 Gaussian noise synthesized by the Box–Muller algorithm (Knuth, 1981) was added in the Gaussian example. The second example was a second-order decay (a single species with an assumed extinction coefficient of unity) with a rate constant of $15\,M^{-1}\,s^{-1}$ and an amplitude of 1. To this was added 0.01 DSX noise using a FORTRAN function DSXDEV given in the Appendix to this paper.

The synthetic data were analyzed with a non-linear fitting program which allowed minimization of either the sum of absolute deviations or the sum of squares. Starting values were provided by the method of successive integration (Matheson, 1987). Data weights of 1 were used for these and all other fits of this work. The results of an NLLAD fit to the Gaussian error case are shown in Fig. 1a. A good fit was obtained, as judged by the coincidence of the data and the fit (lower box). Further evidence of a good fit is shown by the residuals (upper box) which are evenly disposed about zero. The least squares fit, NLLS, was not significantly different from the NLLAD fit as shown in Table 1. The errors shown are 1 mean absolute deviation in the NLLAD case

Table 1

	NLLAD	NLLS
Rate constant	6.02 ± 0.03	6.03 ± 0.04
Amplitude	1.002 ± 0.004	1.003 ± 0.005
Iterations	1	1

and 1 SD in the NLLS case. A satisfactory return of the parameters used for synthesis is obtained in both cases.

A plot of the error distribution histogram for the residuals of Fig. 1a is shown in Fig. 1b. The range is from −3 to 3 SD with a column width of 0.5 SD. The smooth curve is a calculated Gaussian distribution scaled to the two central histogram columns. Although the sum of the absolute deviations was minimized, the error histogram coincides well with the smooth Gaussian curve. The small deviations are probably representative of non-randomness of the random number generator used in the Box–Muller procedure.

Figure 2a shows the results of an NLLAD fit to the DSX error data set. A good fit is obtained as judged by the residuals. NLLAD and NLLS fitting are compared in Table 2.

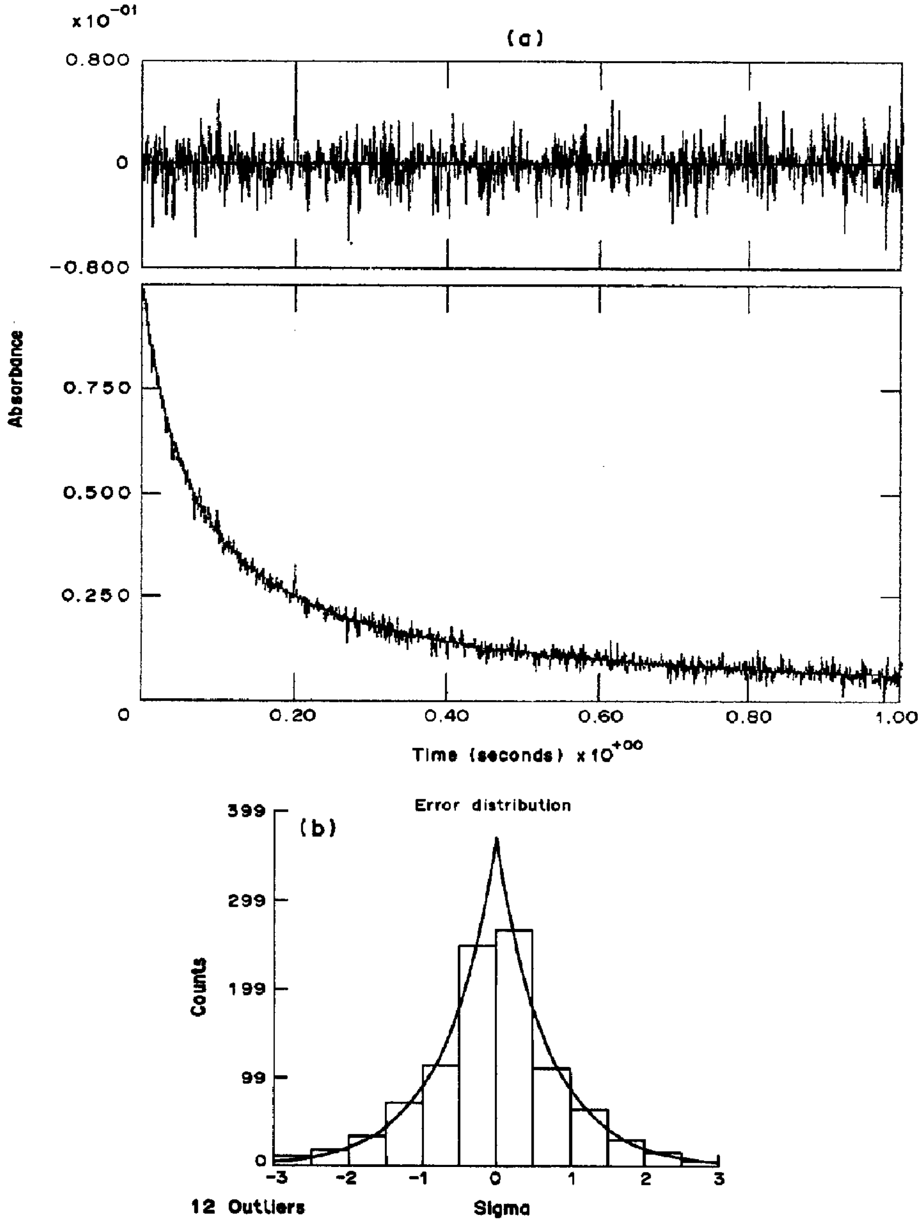


Fig. 2. Synthetic DSX data. Optical density as a function of time (a); histogram of residuals (b).

Table 2

	NLLAD	NLLS
Rate constant	15.03 ± 0.04	15.03 ± 0.06
Amplitude	0.999 ± 0.003	0.999 ± 0.04
Iterations	3	2

As before the NLLAD and NLLS fits return the synthesis parameters well and are not significantly different. The error histogram (Fig. 2b) is clearly non-Gaussian, being relatively sharply-peaked and deficient in the shoulders. A double-sided exponential

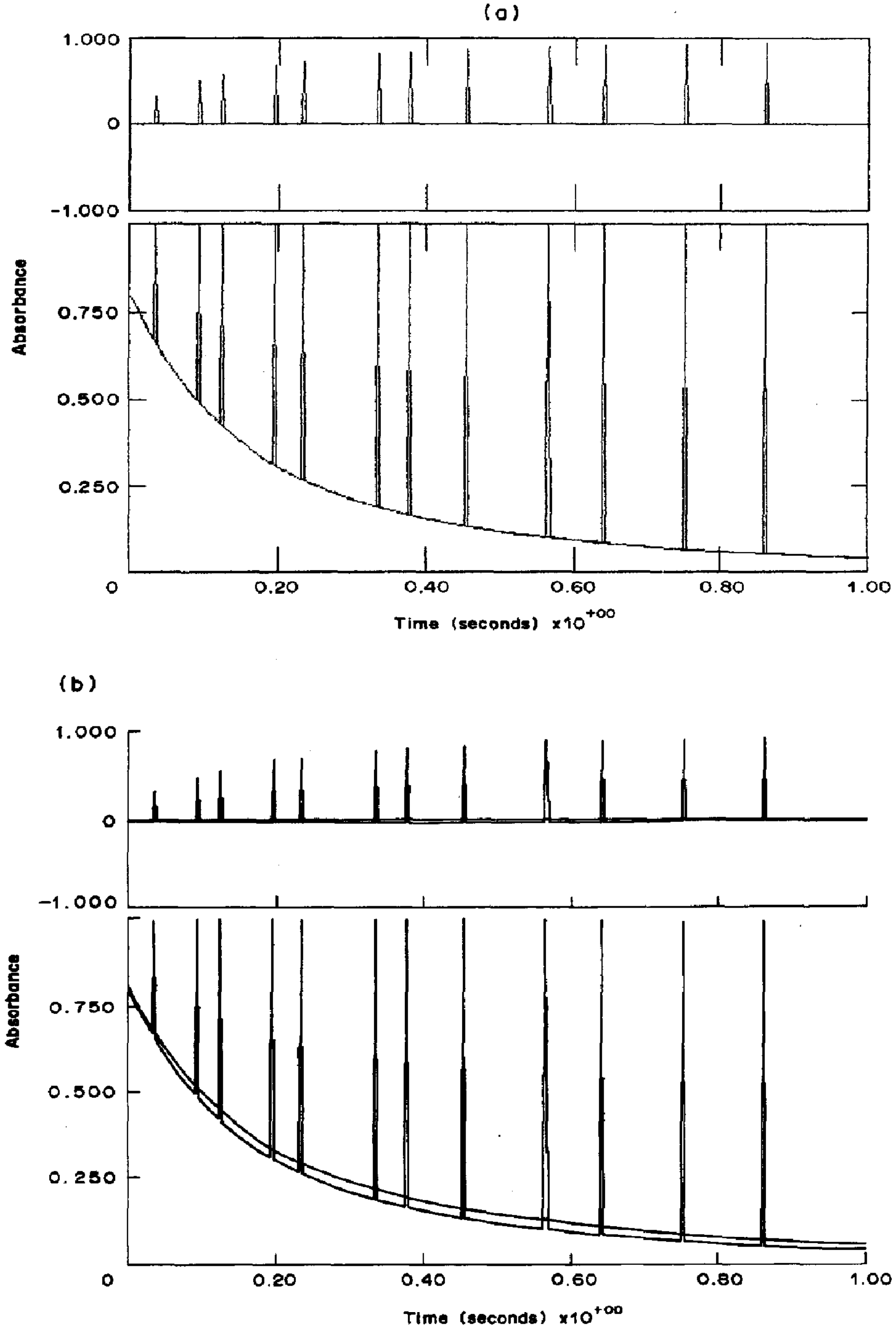


Fig. 3. Synthetic DSX data for two concurrent first-order processes with outliers. NLLAD fit (a); NLLS fit (b).

Table 3

	NLLAD	NLLS
Rate 1	8.0 ± 0.6	8 ± 3
Rate 2	2.0 ± 0.16	1.8 ± 0.8
Amplitude 1	0.49 ± 0.03	0.45 ± 0.18
Amplitude 2	0.30 ± 0.04	0.34 ± 0.19
Iterations	10	1

scaled to the two central columns is shown as a smooth curve.

A third data set was synthesized to simulate the effect of outlying points. The situation considered was a fluorescence decay measured at high sensitivity where occasional cosmic rays were causing very large disturbances. The data set was 400 points for two concurrent first-order processes with rate constants of 8 and $2\ s^{-1}$, and amplitudes of 0.5 and 0.3, respectively. 0.2% of Poissonian noise was added to this. Twelve fullscale defections were further added at random intervals. The results of NLLAD and NLLS fits to this data set are shown in the upper and lower panels of Fig. 3. The NLLAD fit (Fig. 3a) clearly goes through the bulk of the data and is relatively uninfluenced by the outliers. This is not the case for the NLLS fit (Fig. 3b) where the outliers raise the fitted line significantly. The superior performance of NLLAD is also demonstrated by the much better parameter returns shown in Table 3.

Real data

The first case is a stopflow experiment for the reaction of ascorbic acid with dichlorophenol indolephenol (DCIP) at pH 2.5. This may be followed as absorbance loss at 500 nm. The measurements were made on a Kinetic Instruments stopflow machine exhibiting a ringing at early times, possibly due to thermal lensing. The ringing effect is clearly visible in the upper box of Fig. 4a, an NLLAD fit for a first-order decay with a background. As with the synthetic data, a good fit was obtained. The residual histogram (Fig. 4b) is clearly non-Gaussian, and is in fact more sharply peaked than the superimposed DSX curve. NLLAD and NLLS fits are not significantly different, as shown in Table 4.

The second example is a sedimentation equilibrium experiment. The protein was porcine coagulation factor 10, and the measurements were made on a Beckman Prep-Scanner analytical ultracentrifuge. The measured property was optical density at 280 nm, measured as a function of length along the cell x. Only the monomeric form was assumed to be present and the data fitted with the equation of Johnson *et al.* (9):

$$Y = C \cdot \exp(-(x^* x - x_0^* x_0)/2) + b, \qquad (2)$$

with the desired parameter, the molecular weight being deprived from the determined parameters and constants representing the rotor rotation rate, absolute temperature and protein partial molar volume.

NLLAD and NLLS fits to equation (2) are shown in Fig. 5a and b. The residuals (upper boxes) are clearly non-Gaussian in both cases, showing a systemic upward trend at the left end, and pronounced ringing at the right end. The least squares fit has attempted to split the difference of the upward trend at the left end. This effect is much less marked for the least absolute deviation fit. This is to be expected since the residuals contribute as the magnitude to the adjustments in the NLLS case, but only as their sign in the NLLAD case.

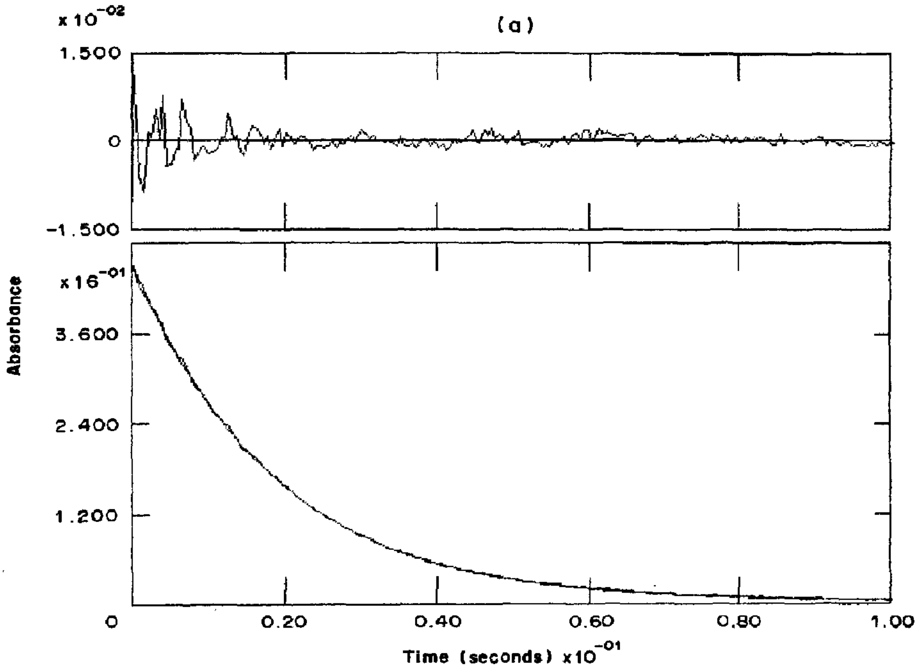
A significant difference is seen in the computed molecular weights $53{,}450 \pm 90$ in the NLLS case and $52{,}470 \pm 50$ in the NLLAD case. The higher value of the least squares case demonstrates the influence of the upward trending outliers at the left end of the residual curve.

A conscientious experimenter would, of course, fit a truncated data set where both the left end upward trend and the marked ringing at the right end are not included. A NLLAD fit to the center part of the ultracentrifuge data set is shown in Fig. 6a. The molecular weight corresponding to the fit is $51{,}000 \pm 40$, and is the most "credible" of the three values given here. The rather complex looking residuals of the upper box are clearly non-Gaussian. This is confirmed by the histogram plot of Fig. 6b which is remarkably coincident with the superimposed smooth DSX curve.

CONCLUSIONS

The results of applying NLLAD and NLLS fits to the synthetic data sets of Figs 1 and 2 show the fits to be equivalent, being equally good at returning the parameters used for synthesis. This good behaviour is shown even though one error distribution is relatively wide and the other sharply peaked, suggesting that in the cases with evenly-distributed errors, the shape of the distribution has little effect on the final converged parameters. This suggests that most least squares fits generate valid parameters, even where the error distribution is non-Gaussian. The data of Fig. 3 with severe outliers are fitted much better by NLLAD than NLLS. This is to be expected since the NLLAD adjustments are proportional to only the sign of the residual as compared to the residual itself for NLLS. Data composed of counts which Fig. 3 simulates are normally analyzed using least squares with Poissonian weights. The malign effect of outliers on the least squares analysis contrasts markedly with the robust analysis. If, for any reason in a counting experiment one or more channels contain spurious counts, then the difference in performance between the two methods is such that given the availability of robust analysis, the use of least squares with Poissonian weighting is probably a pointless exercise.

The real stopflow data of Fig. 4 is equally well fitted by NLLS and NLLAD even though the error



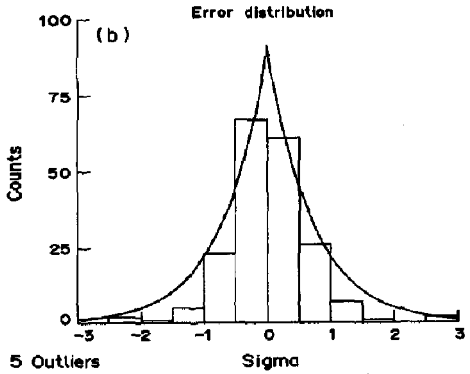


Fig. 4. Real stopflow data. Optical density as a function of time (a); histogram of residuals (b).

histogram, by virtue of the initial ringing, looks more double-sided exponential than Gaussian.

The ultracentrifuge data of Fig. 5 which has residuals unevenly disposed about zero (upper boxes) is much better fitted by NLLAD than NLLS. The double-sided exponential residuals of Fig. 6, where the optical density is between 2 and 3, may give support to the suggestion of Johnson & Frazier (1985) that optical densities cannot be normally distributed. Their argument is that since the optical density is given by the logarithm of a reference transmission to the measured transmission, if the transmission is normally distributed then the optical density cannot be. This is clearly the case for the ultracentrifuge data, but is much less true for stopflow measurements with small optical density changes. The distribution of errors is more nearly Gaussian for such experiments.

The inescapable conclusion is that since the adjustments are only proportional to the sign of the deviations, NLLAD will always be much better for cases where the residuals, and even more strongly the outliers, are not evenly distributed about the function. Since NLLAD performs as well as NLLS on

Table 4

	NLLAD	NLLS
Rate constant	5.438 ± 0.009	5.451 ± 0.015
Amplitude	4.387 ± 0.006	4.393 ± 0.009
Background	−1.73E − 6	−1.74E − 6
Iterations	17	2

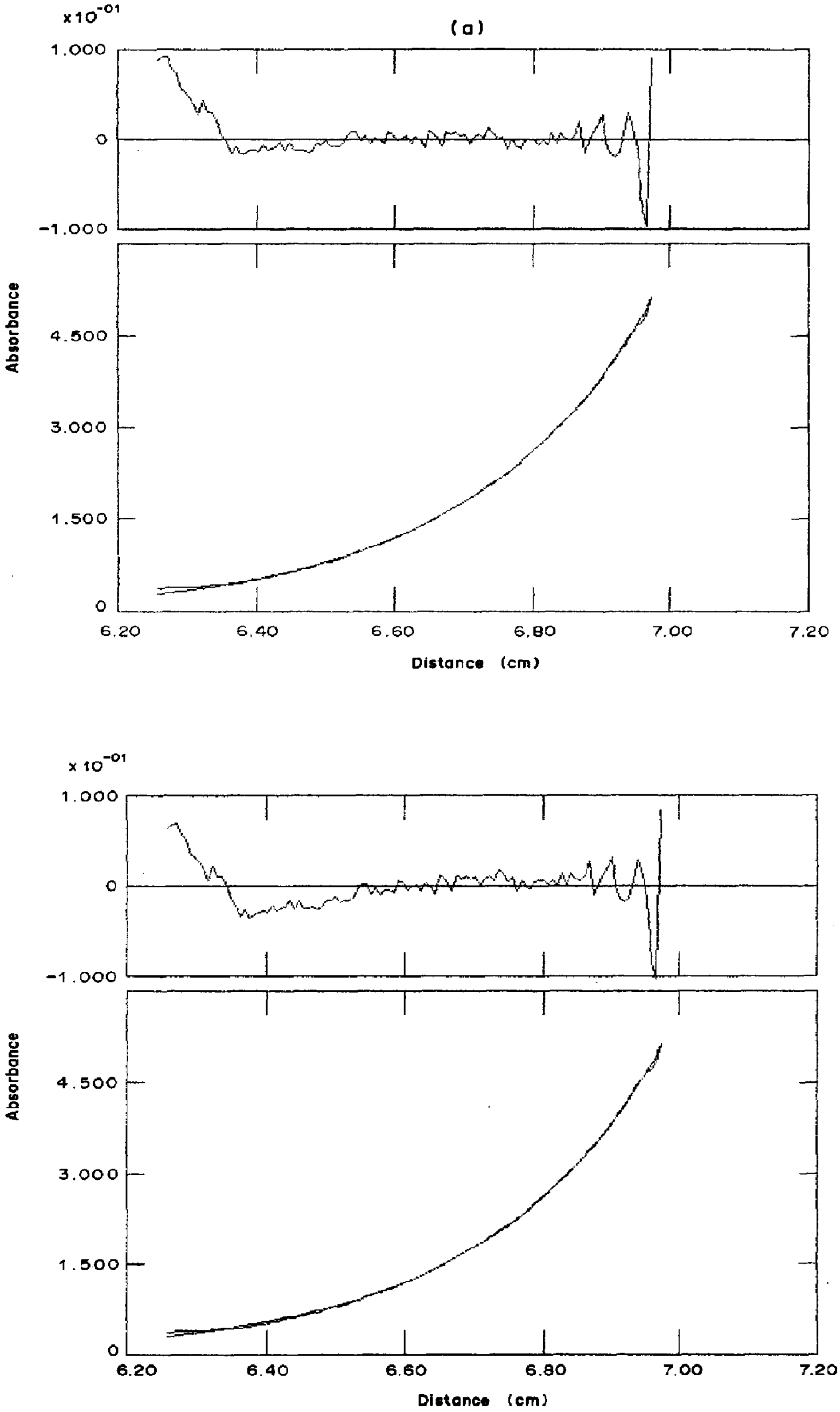


Fig. 5. Ultracentrifuge data, optical density as a function of cell distance. NLLAD fit (a); NLLS fit (b).

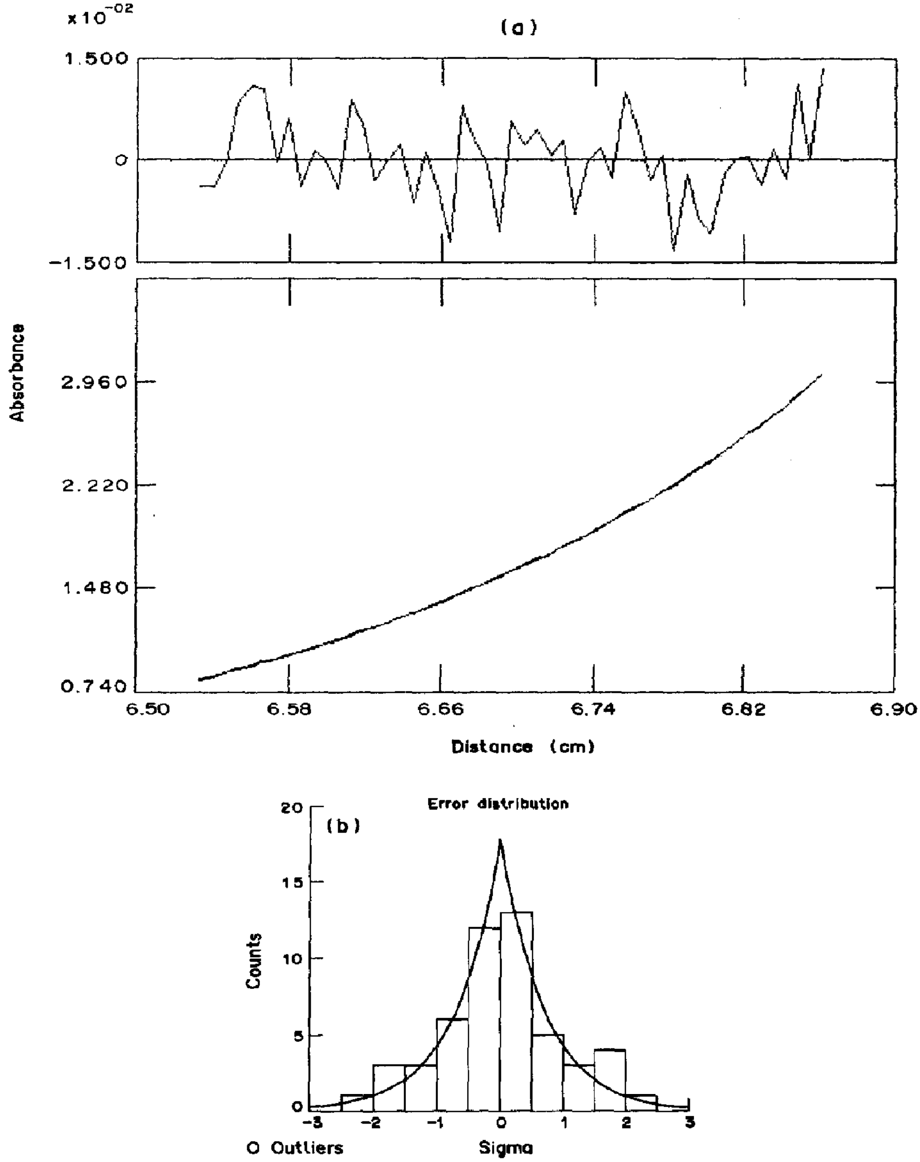


Fig. 6. Truncated data of Fig. 5. NLLAD fit to data (a); histogram of residuals (b).

synthetic Gaussian data and is much better on data with unevenly distributed errors, least absolute deviation fits are *safer* and should be the fitting technique of *first* choice. The only price exacted by the choice of NLLAD over NLLS is that a larger number of iterations may be required for convergence. This may be indicative of a shallower minimum in absolute deviation space than in the variance space of least squares. Given the better performance of NLLAD in unfavourable circumstances and the high floating point performance of modern microcomputers, the resultant small price is well worth paying. If least squares fitting is chosen, the error distribution histogram should be demonstrated to be approximately Gaussian. Where the number of data pairs is too small to justify an error histogram only least absolute deviation fitting should be used, since uneven distributions may be common.

Acknowledgement—The author thanks Whyte G. Owen, Section of Hematology Research, Mayo Clinic, Rochester, Minn. for the ultracentrifuge data and On-Line Instrument Systems for support.

REFERENCES

Bevington P. R. (1969) *Data Reduction and Error Analysis in the Physical Sciences*, p.32 and Chap. 11. McGraw-Hill, New York.

Huber P. (1977) *Robust Statistical Procedures*. Society of Industrial and Applied Mathematics, Philadelphia.
Johnson M. L. & Frazier S. G. (1985) *Methods in Enzymology*, Vol. 117 (Edited by Hirs C. H. W. and Timasheff S. N.) p. 301.
Johnson M. L., Correia J. J., Yphantis D. A., & Halvorson H. R. (1981) *J. Biophys. Soc.* **36**, 575.
Knuth D. E. (1981) *Seminumerical Algorithms*, 2nd Edn, Vol. 2 of *The Art of Computer Programming*, p. 116.
Levenberg K. (1944) *Q. Appl. Math.* **2**, 164.
Marquardt D. R. (1963) *SIAM J. Soc. Ind. Appl. Math.* **11**, 431.
Matheson I. B. C. (1987) *Analyt. Instr.* **16**, 345.
Matheson I. B. C. (1989) *Computers Chem.* **13**, 385
Press W. H., Flannery B. P., Teukolsky S. A. & Vetterling W. T. (1986) *Numerical Recipes*, Chap. 7. Cambridge University Press.

APPENDIX

The Generation of Double-Sided Exponential Deviates

This derivation is given since the double-sided exponential distribution is not well known. Following Press *et al.* (1986), let us assume that we have a random number generator giving uniform deviates between 0 and 1, i.e.

$$p(x) = dx \quad \text{for} \quad 0 < x < 1, \quad p(x) = 0 \quad \text{otherwise.}$$

For a function $y(x)$ the fundamental transformation law of probabilities gives:

$$p(y) = p(x)[dx/dy].$$

If $y(x) = -\ln(x)$ and $p(x)$ is as above then:

$$p(y) = [dx/dy]^* \, dy = \exp(-y)^* \, dy,$$

which is distributed exponentially. The following FORTRAN function assumes that a random number generator generating uniform deviates between 0 and 1, RAND(IDUM), initialized by IDUM, is available as either a system or user-supplied function. Suitable random generators may be found in Press *et al.* (1986). The function DSXSEV yields exponential deviates distributed about zero, i.e. double-sides exponential deviates.

```
      FUNCTION DSXDEV(IDUM)
1     D1 = 2.0*RAND(IDUM) - 1.0
      D2 = ABS(D1)
      IF(D2.EQ.0.0) GOTO 1
      D3 = -LOG(D2)
      DSXDEV = SIGN(D3,D1)
      RETURN
      END
```